Behavioral and neural markers of social vision

Etienne Abassi[1] & Liuba Papeo[1]

[1]*Institut des Sciences Cognitives—Marc Jeannerod, UMR5229, Centre National de la Recherche Scientifique (CNRS) and Université Claude Bernard Lyon 1, 67 Bd. Pinel – 69675 Bron (France)*

Correspondence to:

E.A. etienne.abassi@isc.cnrs.fr

L.P. liuba.papeo@isc.cnrs.fr

Competing interests: All authors declare that they have no financial or non-financial competing interests.

Abstract (199 words)

Research on face perception has revealed highly specialized visual mechanisms such as configural processing, and provided markers of interindividual differences –including disease risks and alterations– in visuoperceptual abilities that traffic in social cognition. Is face perception unique in degree or kind of mechanisms, and in its relevance for social cognition? Combining fMRI and behavioral methods, we address the processing of an uncharted class of socially relevant stimuli: minimal social scenes involving configurations of two bodies spatially close and face-to-face. We report category-specific activity for facing (*vs.* non-facing) two-body configurations in visual cortex. That activity shows face-like signatures of configural processing –i.e., stronger response and greater susceptibility to stimulus inversion for facing (*vs.* non-facing) dyads–, and is predicted by performance-based measures of body-dyad perception. Moreover, individual performance in body-dyad perception is reliable, stable-over-time and correlated with the individual social sensitivity, coarsely captured by the Autism-Spectrum Quotient. Further analyses clarify the relationship between single-body and body-dyad perception. We propose that facing dyads are processed with the same high specialization of the most biologically/socially relevant stimuli: faces. Like face perception, facing-dyad perception can provide a measure of basic (visual) processes that lay the foundations for understanding others, their relationships and interactions.

Keywords: Configural visual processing; Scene perception; fMRI; Individual differences; Autism; Brain-behavior mapping

Introduction (1003 words)

Few functions of the human brain have been explored as much and deeply as the visual processing of faces. And few visual functions appear as vital for social life as face processing. Face processing underlies detection, recognition and identification of conspecifics, and encoding of visual information that provides the basis for attributing states, traits, goals and intentions to others, and regulating one's own behavior in social contexts (Baron-Cohen, 1997; Atkinson et al., 2011). Thus, it is not surprising that highly efficient and specialized mechanisms exist for face perception, in the human visual system.

One of the most robust markers for specialized face processing mechanisms is the *face inversion effect* (FIE), a disproportionate cost in processing a face inverted upside-down relative to a face in the canonical upright orientation (Yin, 1969). The FIE has revealed a visual mechanism, so-called configural, that treats the face not as a set of parts (all present in both upright and inverted stimuli), but as a perceptual unit with a specific configuration (i.e., eyes above mouth), which is disrupted by inversion, yielding catastrophic effects on recognition (Tanaka & Farah, 1993; Maurer et al., 2002).

Besides revealing important aspects of visual functioning, the sensitivity to face inversion is reliable and stable over time at the intraindividual level, and can highlight interindividual differences that capture the normal variability in the general population, as well as disease risks and alterations, in visual perception (Duchaine and Nakayama, 2006; Yovel et al., 2014) and in socially relevant cognitive abilities (Wyer et al., 2012; Neuhaus et al., 2016; Webb et al., 2017). Reduced sensitivity to face inversion –and atypical face processing in general– is common in developmental disorders that affect social cognition such as the autism spectrum disorder (ASD) (Farroni and Senju, 2011; Di Giorgio et al., 2016; Neuhaus et al., 2016; Hadad et al., 2019; Van der Donck et al., 2019; Laycock et al., 2020), although with high variability within the ASD population and no consensus on the neurofunctional basis of the alteration (Brambilla et al., 2003; Johnson, 2004; Dawson et al., 2005; Simmons et al., 2009; Cook et al., 2012).

Is face perception unique in degree or kind of cognitive and neural mechanisms, and in its relevance for social cognition? While faces remain the best understood case of visual specialization, other visual stimuli that support social cognitive tasks, such as bodies and biological motion (Atkinson et al., 2011), may recruit analogous mechanisms (Reed et al., 2006; Troje and Westhoff, 2006). Moreover, recent research indicates an uncharted class of stimuli, whose visual processing might be analogous to faces. Members of this class are two-body shapes, that is, configurations of two bodies spatially close and face-to-face, as is (proto)typical of social interaction (Papeo et al., 2017; Papeo, 2020). Much of what we learn about social life, and how we regulate our social behavior *here and now*, depends on the observation of people interacting with people (Quadflieg and Koldewyn, 2017; Powell and Spelke, 2018). Thus, the visual specialization for face-to-face body dyads (hereafter, facing dyads) would respond to the need for efficient detection and recognition of social interaction in the cluttered and crowded visual world.

Visual processing of facing dyads share with face processing a particular sensitivity to inversion: In fast visual categorization (i.e., discrimination of bodies from other objects), subjects' performance drops dramatically for inverted facing dyads, relative to the same dyads presented upright (Papeo et al., 2017; Papeo and Abassi, 2019). The cost of inversion for facing dyads is significantly larger than the cost of inversion for identical bodies in a non-facing configuration (e.g., back-to-back bodies), an effect known as the two-body inversion effect (2BIE). Interestingly, the cost of inversion for facing dyads is comparable in size to the FIE (see Fig. 1 and Supplementary Information).

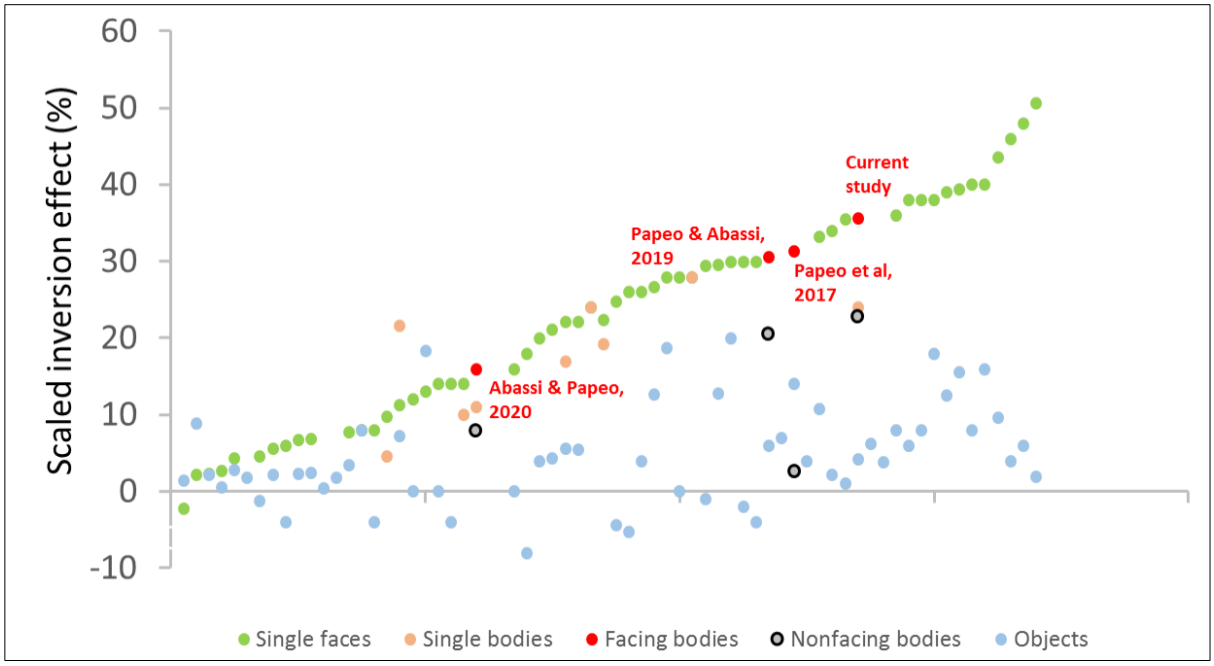

Figure 1. Scaled inversion effect (%) for facing dyads compared to the effect for non-facing dyads, single faces, single bodies and other objects. This figure is based on a meta-analysis in Rezlescu et al. (2016), to which we added values of the scaled inversion effect for facing and non-facing body-dyads, as measured in three previous studies and in the current one. For each study, a scaled inversion effect was computed as: [(raw inversion effect) / (1-chance)]. On average, the size of the inversion effect for facing body dyads is closer to the size of the inversion effect for faces than all other categories. Average scaled inversion effects are (in descending order): 28.4 ±8.55 *SD* for facing bodies (4 studies), 22.78 ±13.78 for faces (55 studies), 18.5 ±7.07 for non-facing bodies (4 studies), 17.71 ±7.74 for single bodies (9 studies), 4.76 ±6.63 for other objects (56 studies). Studies are sorted on the X axis according to the size of the scaled inversion effect for each category in each study. References written in red on the figure correspond to the closest red dot. The full list of studies included in the figure is reported as Supplementary Information (see Table S2).

What is the two-body inversion effect? The present study asked whether, like the FIE, the 2BIE is *a)* an index of visual specialization mediated by configural processing, and *b)* a reliable measure of the individual visuoperceptual functions that traffic in social perception and cognition.

To address the first question, we tested whether activity in body-specific visual areas of the occipitotemporal cortex showed signatures of configural processing during perception of facing dyads. A neural signature of configural face processing is greater activity for faces relative to scrambled or scattered faces (Brandman and Yovel, 2016). Thus, if facing dyads recruit configural processing, they should evoke greater activity in body-specific visual areas, relative to their scattered counterparts, i.e., stimuli featuring the same parts (bodies) in a non-facing configuration. Moreover, the 2BIE captured in the subjects' behavior should predict the activity of body-specific visual areas in response to upright *vs.* inverted dyads, just like the FIE predicts, across subjects, the response to upright *vs.* inverted faces in the face-specific visual cortex (Yovel and Kanwisher, 2005). To test these predictions, we recorded neural activity using functional MRI (fMRI), while healthy subjects viewed upright or inverted facing and non-facing dyads. On a different day, the same subjects performed a visual categorization task to measure the 2BIE (Papeo et al., 2017). We tested the visual specificity of the whole-dyad configuration considering the difference in neural activity for facing *vs.* non-facing dyads, and the correlation between the behavioral 2BIE and the effect of inversion on the neural response to dyads (the neural 2BIE) in individual subjects.

To address the second question, subjects performed the visual categorization task a second time after weeks, so that we could estimate the stability of the 2BIE over time. They also completed the Autism-Spectrum Quotient Test (AQ), a measure of the autism-spectrum traits, sensitive to interindividual differences in the general population, with respect to socially relevant cognitive abilities (Baron-Cohen et al., 2001; Ruzich et al., 2015). Significant correlations between the two measures of the 2BIE (i.e., test-retest reliability) and between the 2BIE and the AQ, at the individual level, will provide first, novel indication that visual perception of social scenes can be a reliable measure of the visuoperceptual functions that contribute to lay the foundations for understanding others relationships and interactions.

Results (2765)

Processing of body dyads

The behavioral 2BIE. To measure the behavioral 2BIE, 22 subjects of the 29 included in the fMRI study (see below), took part in two identical sessions of the same visual categorization task (Papeo et al., 2017). During this task, they saw images of single bodies, facing or non-facing dyads, and of single chairs, and facing or non-facing pairs of chairs. All images were presented upright or inverted upside-down, for a short time (30 ms), and then masked. For each trial, subjects had to report whether they had seen bodies or chairs, regardless of the number of items (one/two), positioning (facing/non-facing) or orientation (upright/inverted). The second session took place on average 20 days after the first session. Accuracy data averaged across the two sessions of the task were analyzed in a 2 Category (body, chair) x 3 Stimulus (facing, non-facing, single stimuli) x 2 orientation (upright, inverted) repeated-measures ANOVA. Results showed a trend for a main effect of Category [$F(1,21) = 3.82$, $p = 0.064$, $\eta_p^2 = 0.15$], significant effects of Stimulus [$F(2,42) = 7.67$, $p = 0.001$, $\eta_p^2 = 0.27$] and Orientation [$F(1,21) = 25.63$, $p < 0.001$, $\eta_p^2 = 0.55$], and significant interactions between Category and Stimulus [$F(2,42) = 8.45$, $p < 0.001$, $\eta_p^2 = 0.29$], Category and Orientation [$F(1,21) = 14.1$, $p = 0.001$, $\eta_p^2 = 0.40$], and Stimulus and Orientation [$F(2,42) = 6.27$, $p = 0.004$, $\eta_p^2 = 0.23$]. Main effects and interactions were qualified by a significant three-way interaction between Category, Stimulus and Orientation [$F(2,42) = 9.13$, $p = 0.001$, $\eta_p^2 = 0.30$]. Two separate ANOVAs with body-trials only and chair-trials only, showed that the magnitude of the inversion effect was affected by the type of Stimulus (single, facing, non-facing), for body-stimuli (Effect of Stimulus: [$F(2,42) = 9.51$, $p < 0.001$, $\eta_p^2 = 0.31$]; Effect of Orientation: [$F(1,21) = 21.25$, $p < 0.001$, $\eta_p^2 = 0.50$], Interaction: [$F(2,42) = 10.78$, $p < 0.001$, $\eta_p^2 = 0.34$]), but not for chair-stimuli (Effect of Stimulus: [$F(2,42) = 3.19$, $p = 0.051$, $\eta_p^2 = 0.13$]; Effect of Orientation: [$F(1,21) = 5.27$, $p = 0.032$, $\eta_p^2 = 0.20$], Interaction: [$F(2,42) = 1.61$, $p = 0.212$, $\eta_p^2 = 0.07$]; see *Supplemental Information* for details on performance with chairs).

Considering more closely the performance with body-stimuli, pairwise *t*-tests (two-tail) showed that, while the inversion effect (i.e., the difference between upright *vs.* inverted stimuli) was significant for all three stimulus conditions [facing dyads: $t(21) = 4.88$, $p < 0.001$; non-facing dyads: $t(21) = 4.28$, $p < 0.001$; single bodies: $t(21) = 4.10$, $p = 0.001$], it was larger for facing dyads, relative to non-facing dyads, $t(21) = 3.53$, $p = 0.002$, and single bodies, $t(21) = 3.76$, $p = 0.001$, and comparable in the last two conditions, $t(21) = 1.21$, $p = 0.240$ (Fig. 2a). Separate analyses for data collected in the first and in the second session of the task gave consistent results (*Supplemental Information* and Fig. S1a). Consistently with previous reports (Abassi and Papeo, 2020; Papeo et al., 2017; Papeo and Abassi, 2019) accuracy proved more sensitive than reactions times (RTs) to the effects of inversion in the current task. RT analyses gave no effects or conformed to the pattern of accuracy results (*Supplemental Information;* Fig. S1b-c).

At the group level, there was no difference in the magnitude of the 2BIE between the first and the second session [$t(21) = 1.13$, $p > 0.250$, two-tail], although accuracy was overall higher in the second session [$t(21) = 2.96$, $p = 0.008$, two-tail], most likely as an effect of learning. The test-retest reliability of the 2BIE was computed considering, for each subject, the two measures of the 2BIE [$(\text{Upright-inverted})_{\text{facing}}$-$(\text{Upright-inverted})_{\text{non-facing}}$], collected in the behavioral session 1 and 2, respectively. We found a significant within-subjects Pearson correlation [$r(20) = 0.63$, $p = 0.002$] , which was significantly higher than any random between-subjects correlation [$p = 0.001$] (Fig. 2b). This result defines the 2BIE as a reliable, stable over time, measure of the individual visual processing of body dyads.

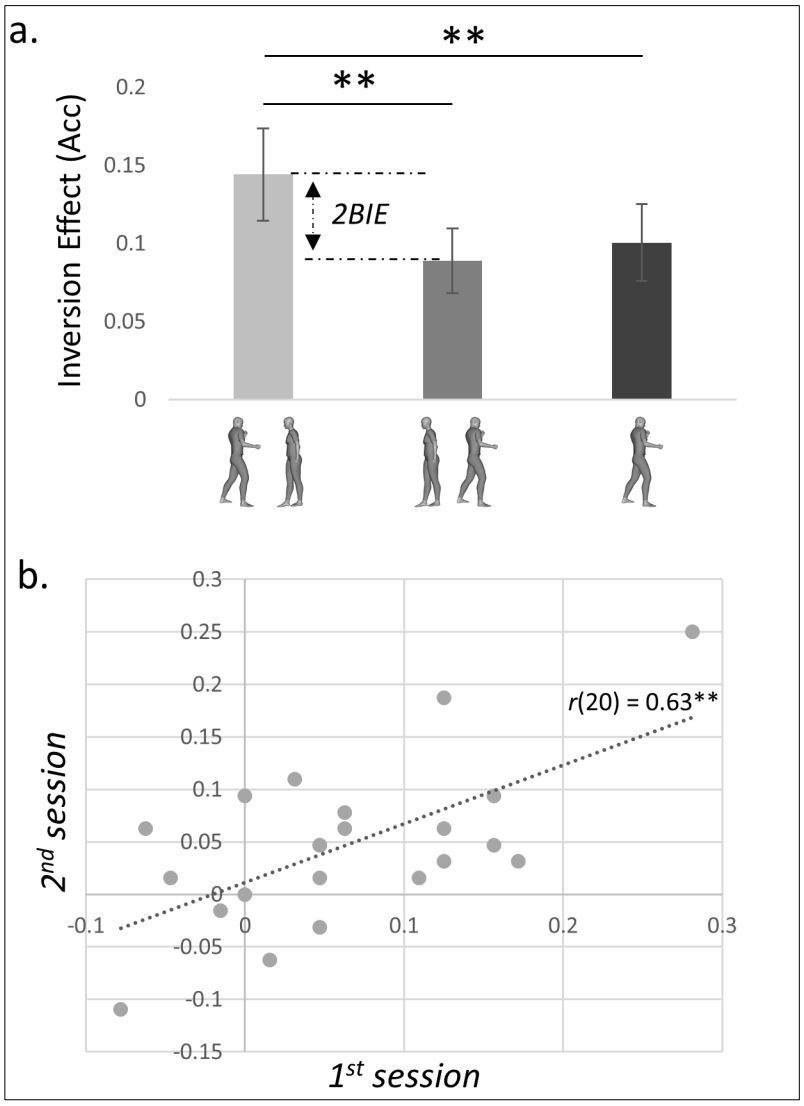

Figure 2. Behavioral 2BIE and within-subject correlation between two measures of the 2BIE. a. Inversion effect (proportion of correct responses for upright minus inverted trials) averaged over the two behavioral sessions for facing dyads, non-facing dyads and single body. Error bars denote the within-subjects normalized standard error of the mean; $^{**}p \leq 0.01$; two-tail. **b.** Pearson correlation between the size of the 2BIE for the same subjects, measured in the first and second session of the visual-categorization task. $^{**}p \leq 0.01$.

Neural signatures of visual configural processing. A neural signature of configural face processing is the increased response to faces (or bodies) *vs.* scattered stimuli in category-specific visual areas of the occipitotemporal cortex (Taylor et al., 2007; Brandman and Yovel, 2016). Here, we report an analogous effect during processing of body dyads, where non-facing dyads were taken as the scattered counterpart of facing-dyad configurations. Converging evidence for increased response to facing *vs.* non-facing dyads was obtained with whole-brain and region of interest (ROI) analyses.

The first analysis was based on data collected during the main fMRI experiment, in which 29 subjects in the fMRI scanner viewed upright or inverted images of facing or non-facing body dyads, and, in independent runs, upright or inverted images of single bodies or objects (see below). The contrast upright facing > non-facing dyads showed an effect in a cluster centered in the left posterior middle temporal gyrus, extending into the lateral occipital and overlapping with the body-specific extrastriate body area (EBA) [MNI peak coordinates: -50 -72 12; peak z-value = 3.96; peak p-value < 0.001; cluster size = 336; p = 0.013, FWE-corrected; see Fig. 3b]. For the contrast inverted facing > non-facing dyads, no clusters survived the FWE correction.

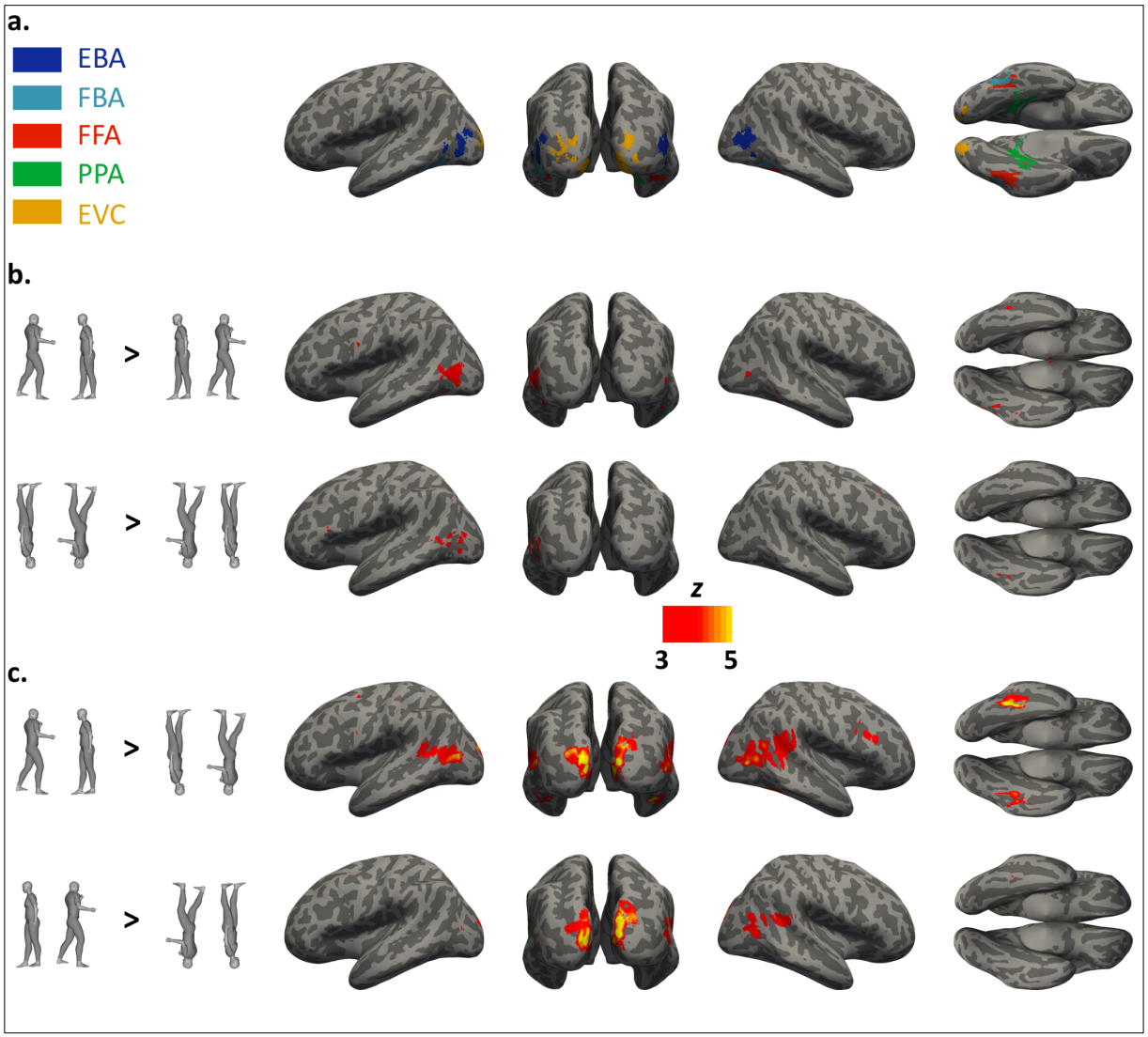

Figure 3. Definition of Regions of Interest (ROIs) and results of whole-brain contrasts addressing the effects of configural processing in body-dyad perception a. ROI location on an inflated brain, based on fMRI data registered during the functional localizer task: EBA (extrastriate body area), FBA (fusiform body area), FFA (fusiform face area), PPA (parahippocampal place area) and EVC (early visual cortex). Group-level clusters are shown for illustration purposes. **b.** Group random-effect maps (n = 29) for the contrast upright facing > non-facing dyads (upper row) and inverted facing > non-facing dyads (lower row). **c.** Group random-effect maps (n = 29) for the contrast upright > inverted facing dyads (upper row) and upright > inverted non-facing dyads (lower row). The color bar indicates uncorrected z values.

The above effect of body positioning in the occipitotemporal cortex was confirmed by the ROI analysis. In particular, using an independent functional localizer task (Fig. 3a) during fMRI, we identified, for

each subject, two areas of the visual occipototemporal cortex with the highest selectivity to body stimuli, the EBA and the fusiform body area (FBA). In addition, we localized the individual face-selective fusiform face area (FFA). The FFA was localized not only because faces were part of our stimuli, but also because this area has consistently been implicated in configural processing of faces as well as bodies (Yovel and Kanwisher, 2005; Brandman and Yovel, 2016). Finally, to control for the specificity of effects in the above ROIs, we defined two additional ROIs in the place-specific parahippocampal place area (PPA), and in the early visual cortex (EVC). From each ROI, we extracted the activity (percentage of signal change) for upright and inverted, facing and non-facing body dyads. A 5 ROIs x 2 Stimulus (facing, non-facing) x 2 Orientation (upright, inverted) ANOVA showed significant effects of ROI [$F(4,112) = 69.89$, $p < 0.001$, $\eta_p^2 = 0.71$], Stimulus [$F(1,28) = 4.85$, $p = 0.036$, $\eta_p^2 = 0.15$] and Orientation [$F(1,28) = 23.92$, $p < 0.001$, $\eta_p^2 = 0.46$], and significant interactions between ROI and Stimulus [$F(4,112) = 12.74$, $p < 0.001$, $\eta_p^2 = 0.31$], and ROI and Orientation [$F(4,112) = 10.65$, $p < 0.001$, $\eta_p^2 = 0.28$], but not between Stimulus and Orientation [$F(1,28) = 1.09$, $p > 0.250$, $\eta_p^2 = 0.03$]. All significant effects and interactions were qualified by a three-way significant interaction [$F(4,112) = 2.69$, $p = 0.035$, $\eta_p^2 = 0.09$], showing that the difference in the magnitude of the inversion effects on neural activity for facing and non-facing dyads changed across ROIs.

Next, we used pairwise *t*-tests to investigate, in each ROI, the two signatures of configural processing targeted here: 1) larger response to facing than to non-facing (upright) dyads, and 2) larger difference between upright and inverted stimuli for facing, than for non-facing dyads (or the neural 2BIE). Given the expected direction of the effects, significance was assessed with one-tailed tests. We found stronger response to upright facing than non-facing dyads in the FFA [$t(28) = 3.41$, $p < 0.001$], EBA [$t(28) = 3.52$, $p < 0.001$] and FBA [$t(28) = 2.51$, $p = 0.009$], but not in the PPA [$t(28) = 0.09$, $p > 0.250$] and EVC [$t(28) = 0.08$, $p > 0.250$] (Fig. 4a). Moreover, we found a larger inversion effect for facing than for non-facing dyads in the FFA [$t(28) = 2.36$, $p = 0.013$], and a trend in the same direction in the EBA [$t(28) = 1.49$, $p = 0.073$]. No other ROI showed a similar effect or trend [FBA: $t(28) = 0.88$, $p = 0.192$; PPA: $t(28) = 0.67$, $p > 0.250$; EVC: $t(28) = 0.74$; $p = 0.234$] (Fig. 4b) (for whole-brain contrasts addressing the effects of inversion, see Fig. 3c; Table S1).

In sum, we found higher activity for facing than non-facing dyads selectively in the EBA, FBA and FFA, that is, in category-specific areas relevant for the processing of the current stimuli. Moreover, neural responses showed a pattern consistent with the behavioral 2BIE (i.e., a larger difference between upright and inverted stimuli for facing than for non-facing dyads) in the FFA and EBA. The relationship between the behavioral and neural BIE was further investigated as follows.

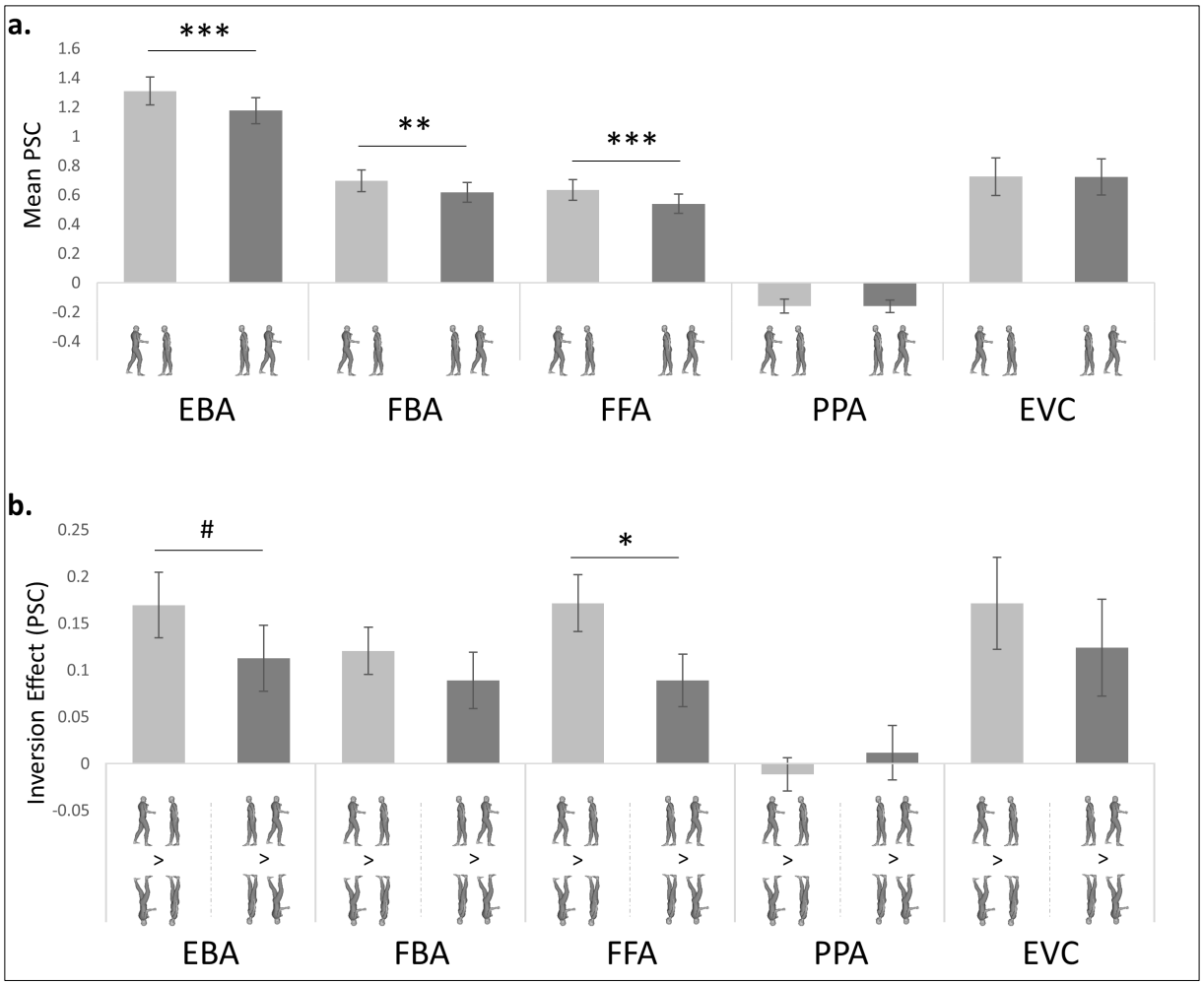

Figure 4. Neural markers of configural processing for body dyads in ROIs. Analyses were performed in the individually defined ROIs. **a.** Activity (percentage of signal change, PSC) for facing and non-facing dyads in each ROI. Higher activity for facing (*vs.* non-facing) dyads is taken as marker of configural processing. **b.** Difference in neural activity (PSC) between upright and inverted stimuli (i.e., inversion effect) for facing and non-facing dyads, in each ROI. Greater inversion effect for facing than non-facing dyads is taken as another marker of configural processing. Error bars denote the within-subjects normalized SEM. $^{\#}p = 0.073$; $^{*}p \leq 0.05$; $^{**}p \leq 0.01$; $^{***}p \leq 0.001$; All tests are one-tail *t*-tests.

Brain-behavior correlation. The current analysis sought to investigate whether the behavioral 2BIE, taken as a signature of the configural processing of dyads, could predict the response to dyadic stimuli in visual cortex. This analysis was carried out within the ROIs and across the whole-brain. First, for each subject who were included in all the behavioral and fMRI sessions of the study (N = 22), for each ROI, we computed a measure of the neural 2BIE as $[(\text{Upright-inverted})_{\text{facing}}/(\text{Upright+Inverted})]–[(\text{Upright-inverted})_{\text{non-facing}}/(\text{Upright+Inverted})]$. Normalized measure of activity was used to take care of inter-subjects differences in the overall magnitude of fMRI response, which are not suitable for correlation with behavioral responses (Yovel and Kanwisher, 2005). Pearson correlations were computed between the individual neural and behavioral 2BIE (accuracy data), revealing a significant effect in the bilateral EBA [$r(20) = 0.44$, $p = 0.039$], but not in other ROIs [FBA: $r(20) = 0.16$, $p > 0.250$; FFA: $r(20) = 0.28$, $p = 0.214$; PPA: $r(20) = -0.24$, $p > 0.250$; EVC: $r(20) = 0.27$, $p = 0.228$] (Fig. 5a). For this analysis, we considered the subjects' performance in the first session of the behavioral task, as that was devoid of learning effects. The same analysis using data from the second session of the behavioral task, showed a consistent, significant effect in the right EBA (see *Supplemental Information* for details in correlation

analysis using data from the second session, average performance across the two sessions, and for separate left and right ROIs).

A separate data-driven analysis provided converging evidence for a 2BIE in the EBA. For each subject, for each voxel that responded to body stimuli across the whole brain (*Methods*), we computed the neural 2BIE and performed the correlation with the behavioral 2BIE. A significant brain-behavior correlation was found in a bilateral cluster in the lateral occipital cortex (Fig. 5b), encompassing the EBA and extending more posteriorly into the occipital cortex [(right peak MNI coordinates: 44 -82 -6; peak *z*-value = 3.43; peak *p*-value < 0.001; cluster size = 219), (left peak MNI coordinates: -48 -74 2; peak *z*-value = 2.83; peak *p*-value = 0.002; cluster size = 52)]. The effect was the strongest in the right hemisphere, which survived correction for multiple comparisons (see *Supplemental Information*).

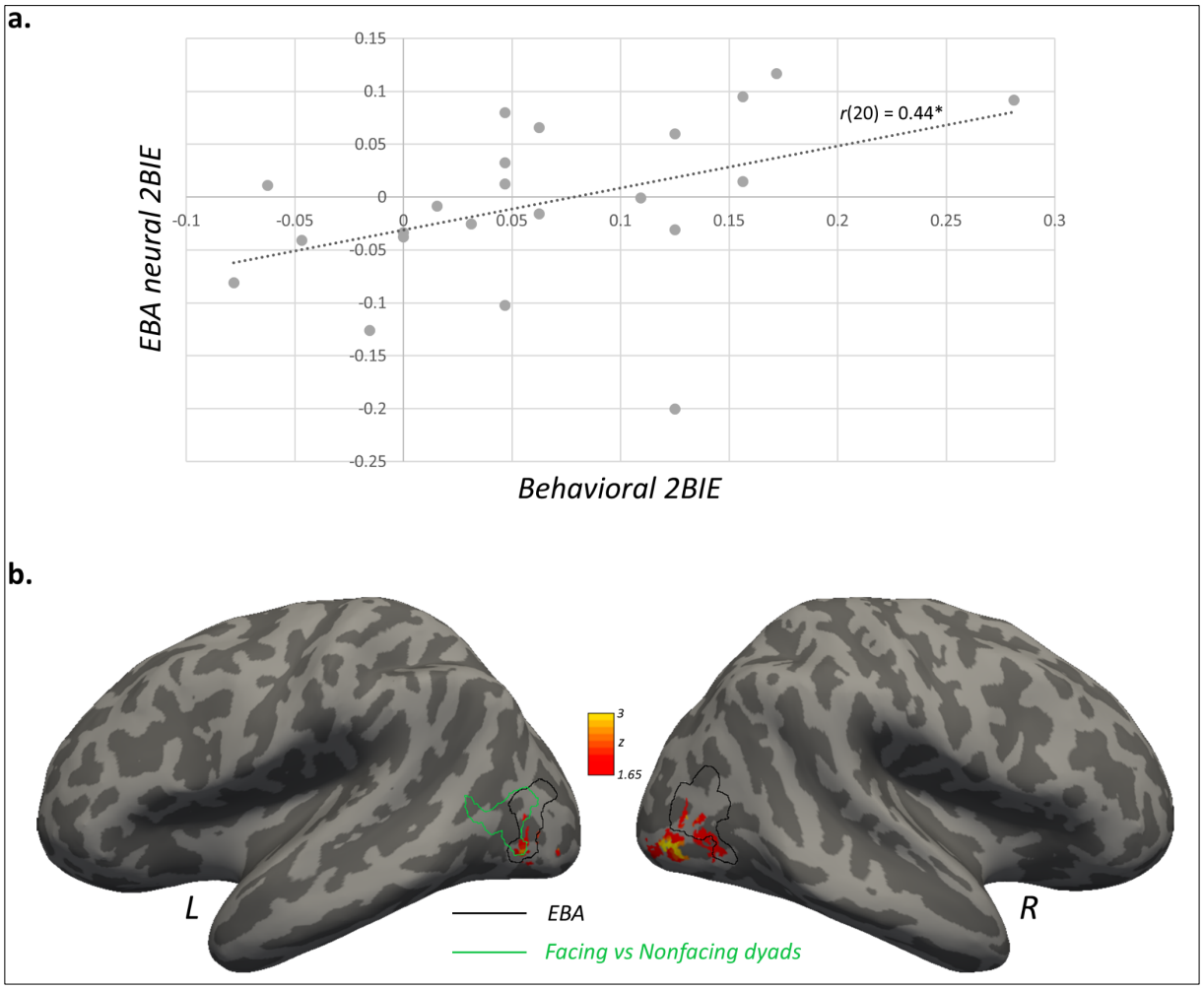

Figure 5. Brain-behavior mapping for the 2BIE. a. Correlation between the behavioral 2BIE (average performance in session 1 and 2 of the visual-categorization task) and the neural-2BIE in the functionally defined EBA. *$p \leq 0.05$ (one-tail). **B.** Statistical map (uncorrected) representing the results of the correlation between the neural 2BIE and the behavioral 2BIE computed in each body-responsive voxel across the whole brain. The color bar indicates *z* values. The region delimited by the black line corresponds to the EBA found with the group-level contrast bodies > [objects+faces+places] using data recorded during the functional localizer task. The region delimited by the green line corresponds to the group-level cluster found with the contrast facing > non-facing dyads in the analysis of data recorded during the main fMRI experiment.

Correlation between the 2BIE and AQ. The above results show that the behavioral 2BIE not only predicts the processing of body dyads in the visual cortex, but also provides a reliable measure of the individual processing of social (multiple-person) scenes, which may be useful to capture interindividual differences. Sensitivity to inversion has been shown to be predictive of the individual's social abilities (e.g., Wyer et al., 2012). To study whether a similar relationship can be established for the 2BIE, all participants in the two behavioral sessions of the fMRI study also completed the Autism Quotient test (AQ), a self-report measure of autistic traits, widely used to quantify the individual's socially relevant cognitive abilities (Baron-Cohen et al., 2001). Correlation between the individual AQ scores and behavioral 2BIE was significant and negative (Spearman *rho*(20) = -0.45; $p = 0.035$), showing lower sensitivity to dyad inversion in individuals with higher scores for autistic traits (see *Supplemental Information* for details in correlation analysis using data from the second session and average performance across the two sessions)

Processing of single bodies

The body inversion effect. So far, we have pushed an analogy between facing-dyad and face processing. Configural processing are also thought to apply to single bodies, although the underlying mechanisms are, for the most part, functionally and anatomically distinct from face-related processing (Yovel and Kanwisher, 2004; Peelen et al., 2006; Peelen and Downing, 2007), and, in many aspects, less clear (Brandman and Yovel, 2012, 2016). To contribute to this issue, we submitted the processing of single bodies to the tests of configural processing applied above, in dyad perception. Using fMRI data from runs involving single bodies and non-body objects, we extracted the neural response to upright and inverted stimuli from each ROI. As a signature of the BIE, we tested whether the difference in the response to upright bodies *vs.* inverted ones was statistically larger than a general effect of inversion that would be observed for any (non-body) stimulus. We ran separate 2 Category x 2 Orientation ANOVAs for each ROI. In the EBA, we found a stronger response to bodies than objects [$F(1,28) = 160.23$, $p < 0.001$, $\eta_p^2 = 0.85$], and to upright than inverted stimuli [$F(1,28) = 7.81$, $p = 0.009$, $\eta_p^2 = 0.22$], but no interaction between Category and Orientation [$F(1,28) = 0.47$, $p > 0.250$, $\eta_p^2 = 0.02$]. Although the interaction was not significant, pairwise *t*-tests showed greater activity for upright *vs.* inverted bodies [$t(28) = 2.15$, $p = 0.02$; one-tail], and a trend for upright *vs.* inverted objects [$t(28) = 1.56$, $p = 0.065$; one-tail]. Identical results were found in the FBA [effect of Category: $F(1,28) = 38.30$, $p < 0.001$, $\eta_p^2 = 0.58$; effect of Orientation: $F(1,28) = 7.53$, $p = 0.010$, $\eta_p^2 = 0.21$; Interaction: $F(1,28) = 0.24$, $p > 0.250$, $\eta_p^2 = 0.01$; upright *vs.* inverted bodies: $t(28) = 2.30$, $p = 0.015$, one-tail; upright *vs.* inverted objects: $t(28) = 1.45$, $p = 0.079$, one-tail] and in the FFA [effect of Category: $F(1,28) = 41.26$, $p < 0.001$, $\eta_p^2 = 0.60$; effect of Orientation: $F(1,28) = 23.59$, $p < 0.001$, $\eta_p^2 = 0.46$; Interaction: $F(1,28) = 1.30$, $p > 0.250$, $\eta_p^2 = 0.04$; upright *vs.* inverted bodies: $t(28) = 4.21$, $p < 0.001$, one-tail; upright *vs.* inverted objects: $t(28) = 2.30$, $p = 0.015$, one-tail]. In the PPA, activity was stronger for objects than for bodies [$F(1,28) = 46.48$, $p < 0.001$, $\eta_p^2 = 0.62$], but there was no effect of orientation, [$F(1,28) = 0.99$, $p > 0.250$, $\eta_p^2 = 0.03$], or interaction [$F(1,28) = 1.93$, $p = 0.176$, $\eta_p^2 = 0.06$], no difference between upright and inverted bodies [$t(28) = 0.29$, $p > 0.250$, one-tail], but stronger response for upright *vs* inverted objects [$t(28) = 1.71$, $p = 0.049$, one-tail]. In the EVC, we found a trend for stronger response to objects than bodies [$F(1,28) = 4.05$, $p = 0.054$, $\eta_p^2 = 0.13$], and a stronger response to upright than inverted stimuli [$F(1,28) = 7.56$, $p = 0.010$, $\eta_p^2 = 0.2$], but no interaction [$F(1,28) = 0.19$, $p > 0.250$, $\eta_p^2 = 0.1$], and a stronger response to upright *vs.* inverted bodies [$t(28) = 1.99$, $p = 0.028$, one-tail] and to upright *vs.* inverted objects [$t(28) = 2.09$, $p = 0.023$, one-tail]. In sum, activity was stronger for upright than inverted bodies in all body- and face-specific ROIs considered here (Fig. 6a), but in no ROI the BIE was statistically different from the effect of inversion on the neural response to non-body objects.

Using another measure of configural body processing, the discrimination accuracy for upright *vs.* inverted bodies in multivariate pattern analysis (MVPA) (see Brandman and Yovel, 2016), we found more compelling evidence for a BIE in face- and body-specific ROIs. This analysis showed a better decoding of upright *vs.* inverted stimuli for both body and non-body objects in the EBA, FBA, FFA, EVC, and only for non-body objects in the PPA (Fig. 6b; Table 1). It also showed that discrimination of multivariate patterns representing upright *vs.* inverted stimuli was more accurate for bodies than for non-body objects in the EBA, FBA and FFA, but not in the PPA and EVC [EBA: $t(28) = 6.17$, $p < 0.001$; FBA: $t(28) = 3.20$, $p = 0.003$; FFA: $t(28) = 2.05$, $p = 0.050$; PPA: $t(28) = 1.90$, $p = 0.067$; EVC: $t(28) = 0.85$, $p > 0.250$]. However, correlations between the individual behavioral BIE and the neural BIE yielded no significant effect in any ROI [EBA: $r(20) = -0.10$, $p > 0.250$; FBA: $r(20) = -0.05$, $p > 0.250$; FFA: $r(20) = -0.12$, $p > 0.250$; PPA: $r(20) = 0.88$, $p > 0.250$; EVC: $r(20) = 0.07$, $p > 0.250$], or in any voxel responsive to bodies, across the brain.

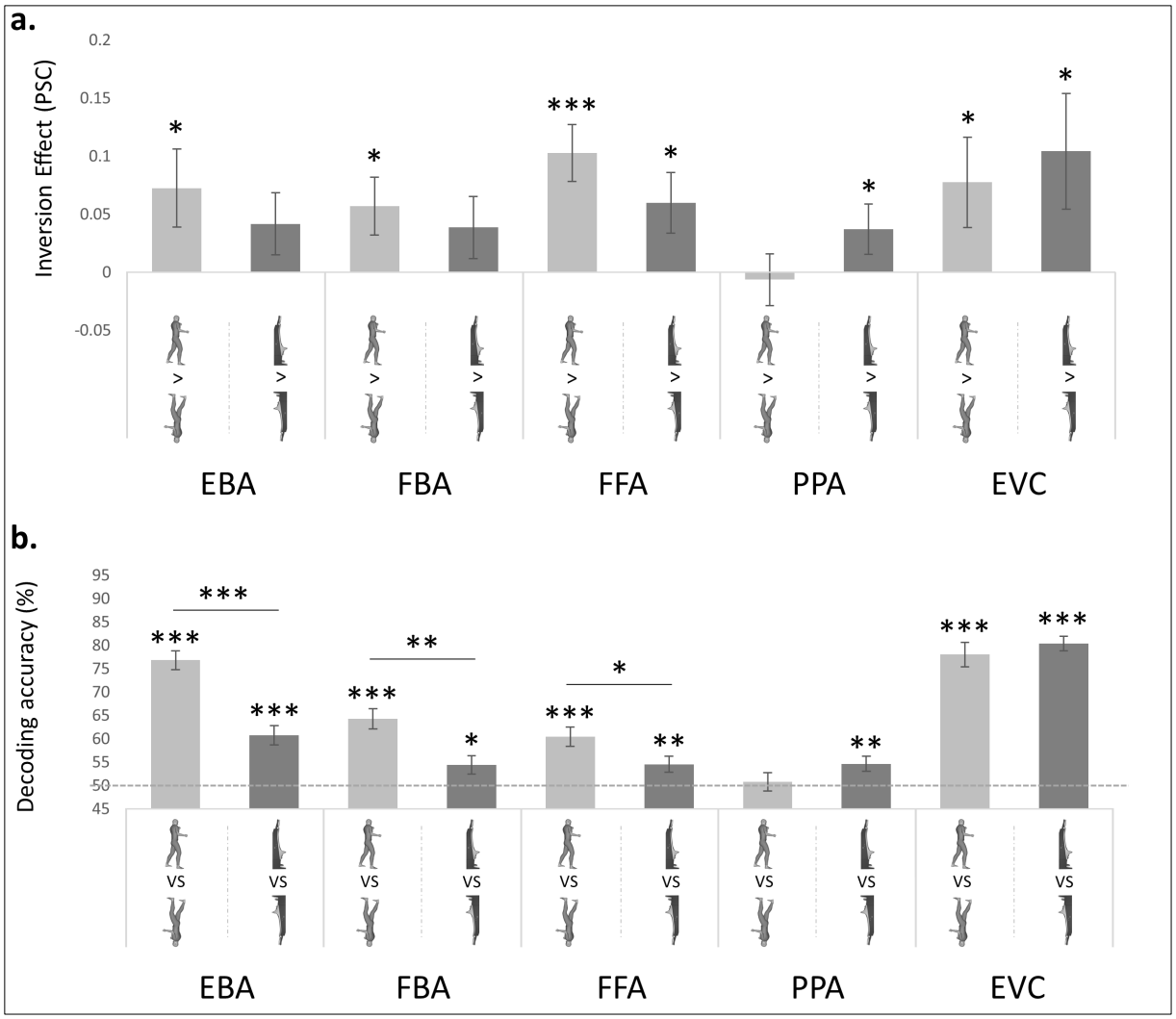

Figure 6. Neural markers of configural body processing in ROIs. a. Difference in neural activity (PSC) for upright *vs.* inverted stimuli (i.e., inversion effect) for bodies and objects, in each ROI. In no ROI, the inversion effect was larger for bodies than objects. $*p \leq 0.05$; $***p \leq 0.001$ All tests are one-tail *t*-tests. **b.** Classification accuracies for decoding of upright *vs.* inverted bodies and upright *vs.* inverted objects, in each ROI. Higher deciding accuracy for bodies than objects is taken as a marker of configural processing. Error bars denote the within-subjects normalized SEM. $*p \leq 0.05$; $**p \leq 0.01$; $***p \leq 0.001$; Comparisons against chance level (50 %) are one-tail *t*-tests. Comparisons of accuracies between bodies and objects are two-tail *t*-tests.

	Bodies			Objects		
ROI	***Accuracy***	***t (28)***	***p-value***	***Accuracy***	***t (28)***	***p-value***
EBA	76.80 (± 2.04)	13.16	*** < 0.001	60.70 (±2.09)	5.12	*** < 0.001
FBA	64.22 (± 2.15)	6.61	*** < 0.001	54.38 (± 1.97)	2.23	** 0.017
FFA	60.42 (± 2.02)	5.15	*** < 0.001	54.53 (±1.72)	2.63	** 0.007
PPA	50.72 (± 1.95)	0.37	> 0.250	54.60 (± 1.61)	2.85	** 0.004
EVC	78.02 (± 2.62)	10.70	*** < 0.001	80.39 (± 1.56)	19.44	*** < 0.001

Table 1. Mean classification accuracy (± SEM) and significance for the decoding of upright *vs.* inverted single bodies and upright *vs.* inverted single objects as a multivariate test of the inversion effect. $^{**}p \leq 0.01$; $^{***}p \leq 0.001$; All tests are one-tail *t*-tests

Discussion (1377)

Our findings show that, during processing of multiple bodies, face- and body-specific visual areas encode spatial relations between bodies, yielding different neural responses to face-to-face and back-to-back bodies. In particular, neural responses to facing bodies show signatures of configural processing, such as stronger activity relative to the response to their scattered version (non-facing dyads), greater susceptibility to stimulus inversion, and correlation with the behavioral 2BIE, proposed as a marker of configural facing-dyad processing (Papeo et al., 2017). Finally, we found that the 2BIE is reliable at the individual level, across behavioral sessions and across brain and behavioral measurement, and correlates with the individual's social sensitivity coarsely captured by the AQ.

Configural processing of body dyads. Previous studies have reported increased activity for facing than non-facing dyads in face- and body-specific visual areas (Walbrin and Koldewyn, 2019; Abassi and Papeo, 2020; Bellot et al., 2021). Thus, depending on their relative spatial positioning, the same two bodies are visually processed in different ways. This result is compatible with growing evidence that the processing of spatial relations is part of, and affects, the object recognition process (Kaiser et al., 2019; Abassi and Papeo, 2020; Papeo, 2020; Hafri and Firestone, 2021). In addition, here we showed particularly high sensitivity of the EBA and FFA to the inversion of facing (*vs.* non-facing) dyads, which characterizes the increased neural response to facing dyads, in terms of configural processing.

Although there was only a trend for the interaction between Stimulus (facing/non-facing dyads) and Orientation (upright/inverted) in the EBA, this was the only ROI, in which the activity was predicted by the 2BIE, as measured during the visual-categorization task. This result was confirmed by the whole brain-behavior correlation, which tested every voxel responsive to body-stimuli across the brain, and revealed an effect in the lateral occipital cortex. We propose that this effect, encompassing the EBA and peaking in an occipital site between the EBA and a posterior occipital area specialized in face processing (occipital face area; Gauthier et al., 2000), reveals a new uncharted area of functional specialization for processing spatial relations between bodies, or people.

Like the EBA (in effect, more strongly than the EBA), the FFA showed high sensitivity to the inversion of facing (relative to non-facing) dyads (i.e., Stimulus by Orientation interaction), but no correlation with the behavioral 2BIE. One possibility is that the FFA response is tied to the processing of faces (and their spatial relations), which were clearly visible during fMRI, and less visible, or less attended to, in the behavioral task (in the visual-categorization task, subjects were instructed to recognize bodies, which varied in posture but always had the same sketchy head/face; see Fig. 2). A version of the visual-categorization task with more emphasis on faces and face processing might yield brain-behavioral correlation in the FFA. This reasoning implies that the visual processing of spatial relations would be to some extent category-specific, involving the same areas that respond preferentially to the object class

implicated in the relation: body-specific areas for bodies, and face-specific areas for faces. However, the recruitment of the FFA in a task that emphasizes body processing is also compatible with the view that the FFA is involved in configural analysis of both faces and bodies (Brandman and Yovel, 2016; see also Gauthier et al., 1999; Rhodes et al., 2004). On this view, different response profiles in the FFA and EBA may reflect the implementation of different aspects, or kinds, of configural processing (Yovel et al., 2014; Rezlescu et al., 2017).

2BIE and interindividual differences. Face recognition abilities vary across individuals (Dennett et al., 2012; Susilo et al., 2013; Yovel et al., 2014; Wang et al., 2012), but are fairly consistent at the individual level (Rule et al., 2009; Lander et al., 2018; Bate et al., 2019; Ramon et al., 2019), and best predicted by the magnitude of face inversion effect, among the markers of configural face processing (Rezlescu et al., 2017). Evidence for reduced sensitivity to face inversion in ASD subjects has encouraged the thinking that the face inversion effect can be used as a marker of alterations and disease risks that impact social cognition (Farroni and Senju, 2011; Di Giorgio et al., 2016; Neuhaus et al., 2016; Hadad et al., 2019; Van der Donck et al., 2019; Laycock et al., 2020).

Along those lines, we showed that, in body-dyad perception, the 2BIE varies across individuals, is reliable over time at the individual level, and correlates with a physiological measure of the same processing (i.e., activity in visual areas for face/body processing; see Elliott et al., 2020). Moreover, we found that the 2BIE correlates with autism-spectrum traits, as measured with the AQ test, so that the magnitude of the 2BIE decreases as the AQ increases. The relationship between visual configural processing for body-dyads –as well as for faces and biological motion (Wyer et al., 2012; van Boxtel et al., 2017)– and the set of abilities measured with the AQ, remains unclear. While we work toward explaining this relationship, effects that reliably characterize the individual's performance, highlighting population-level and interindividual differences, can offer tools for detection of diseases or vulnerabilities. In this spirit, our results encourage further investigation on the clinical significance of the (behavioral and neural) 2BIE as a biomarker of socially relevant visuoperceptual functions.

Body dyads and single bodies. Targeted as measure of configural processing, the body inversion effect (BIE) has been demonstrated with behavioral tasks analogous to those used in face perception (Reed et al., 2003; 2006), and effects in the electroencephalography signal (Stekelenburg and De Gelder, 2004; Orlandi and Proverbio, 2020; Adibpour et al., 2021). However, different fMRI measures of the BIE (e.g., neural adaptation or MVPA) have yielded only inconsistent effects of body inversion in the FFA and/or in the EBA (Brandman and Yovel, 2012, 2016), and some studies have ascribed those effects to the processing of head/face information rather than whole body (Brandman and Yovel, 2010, 2012). Thus, it remains debated whether bodies evoke reliable inversion effect, or recruit configural processing at all.

Our results confirm the ambiguity of the BIE. The behavioral measure of the BIE showed a cost of inversion that was smaller than the cost for facing dyads (and comparable to the cost for non-facing dyads; Fig. 2). Moreover, the effect of body inversion in body- and face-specific visual areas was statistically different from a general effect of inversion (i.e., inversion of non-body objects) in the multivariate test (discrimination of upright *vs.* inverted stimuli), but not in the univariate test (level of activity for upright minus inverted bodies) and, in no ROI, it correlated with the behavioral BIE.

While leaving open the question of the origin of the BIE and its relationship with the 2BIE, our findings suggest that the two effects differ in kind and/or degree of underlying mechanisms. Moreover, we note that the behavioral 2BIE was correlated with neural activity that partly –but not entirely– overlapped with the EBA, as defined with the classic contrast [single bodies > non-body objects]. This observation is in line with the hypothesis that the visual processing of body dyads would recruit additional or partly different mechanisms (and neural structures) with respect to the visual processing of (single) bodies.

Conclusions

Using a multimodal approach, we connected behavioral (the 2BIE) and neural phenomena (greater sensitivity to inversion for facing *vs.* non-facing dyads and larger response to facing *vs.* non-facing dyads), which together demonstrate the recruitment of configural processing in visual perception of minimal social scenes (i.e., facing-body dyads), analogously to what is described for face perception. The recruitment of configural processing suggests that facing dyads are processed with the same efficiency and high specialization of other highly familiar or biologically relevant classes of stimuli, such as faces, bodies and biological motion. Our investigation on the 2BIE also shows that this effect matches –i.e., can account for– the neural effects of body dyad perception, is reliable at the individual level, and correlates with the individual's AQ. The current characterization of the 2BIE thus supports its value as a marker of configural processing and, possibly, a biomarker of individual visuoperceptual abilities. The visuoperceptual abilities captured by the 2BIE could be precisely those that lay the foundations for living in a social world, and for our ability to understand others' relationships and interactions.

Material and Methods

Participants. Thirty subjects took part in the fMRI study (16 female; mean age 24.8 years, *SD* = 4.6). All had normal or corrected-to-normal vision and reported no history of psychiatric or neurological disorders, or use of psychoactive medications. They were screened for contraindications to fMRI and gave informed consent before participation. Of the 30 subjects who enrolled in the fMRI experiment, 23 agreed to come back twice, to take part in two sessions of the behavioral study (13 females; mean age 24.6 years ± 4.5 *SD*). This number was comparable to the sample size used in previous research that used similar task and design (Papeo et al., 2017). From behavioral and fMRI data analyses, we excluded one subject for whom average performance in the behavioral task was >3 *SD* above the group mean. Thus, the final fMRI analyses included 29 participants, and the final behavioral analyses included 22 participants. This study was approved by the local ethics committee (CPP Sud Est V, CHU de Grenoble).

Stimuli

Behavioral study. Sixteen grayscale renderings of single human bodies in profile view and various biomechanically possible poses, were created and edited with Daz3D (Daz Productions, Salt Lake City) and the Image Processing Toolbox of MATLAB (The MathWorks Inc, Natick, Massachusetts). As many bodies were obtained by flipping horizontally each unique body, which yielded to a total of 32 single bodies. Sixteen facing dyads were created from the 32 single bodies, and then they were horizontally flipped to create 16 new dyads, for a total of 32 facing dyads. Non-facing dyads were created by swapping the position of the two bodies in each facing dyad (i.e., the body on the left side was moved to the right side and *vice versa*). The distance between bodies (i.e., the distance between the two closest points of the two bodies in a dyad) was matched across facing and non-facing stimuli ($mean_{facing}$ = 82.88 pixels ± 13.76 *SD*; $mean_{non\text{-}facing}$ = 83.06 ± 13.86 *SD*; $t(15) = 1.00$; $p > 0.250$). For single-body stimuli, the center of image corresponded to the center of the body, defined as the mid-point between the most extreme point on the left and on the right along the x-axis. For dyads, the center of the image was between, at equal distance from, the centers of the two bounding boxes that contained each body. In summary, the final stimulus set included 96 stimuli: 32 single bodies, 32 facing dyads and 32 non-facing dyads, differing only for the relative spatial positioning of the two bodies. Another set of stimuli included 96 images of single chairs (32) and pairs of chairs (32 facing and 32 non-facing). Chair-stimuli were created from 16 grayscale exemplars of chairs and their horizontally flipped version, which were

combined in 16 pairs of facing pairs and their flipped version, and 16 pairs of non-facing chairs and their flipped version. All body- and chair-stimuli were inverted upside-down, yielding a total of 384 stimuli presented against a light grey background. The same number of masking stimuli was created, consisting of high-contrast Mondrian arrays (11° x 10°) of grayscale circles of variable diameter (0.4°-1.8°).

fMRI study. The fMRI study involved the 192 (upright and inverted) body-stimuli used in the behavioral study (64 single bodies, 64 facing dyads and 64 non-facing dyads), in addition to the same 32 single chairs and a new set of 32 machines. Images in the last set depicted one of eight electronic devices (four automated teller machine, and four game machines; i.e., slot machines or arcade game machines) presented in lateral view upright (N=8) and inverted (N=8), and their horizontally flipped version, upright (N=8) and inverted (N=8).

Procedures

Visual categorization task. The task included two identical runs, each containing 32 trials for each of the twelve conditions (upright and inverted single, facing and non-facing, bodies and chairs), presented in random order. Each stimulus appeared once in a run. Subjects sat on a chair, 60 cm away from a computer screen, with their eyes aligned to the center of the screen (17-in. CRT monitor; 1024x768 pixel resolution; 60-Hz refresh rate). Stimuli on the screen did not exceed 7° of visual angle. Each trial included the following sequence of events: blank screen (200 ms), fixation cross (500 ms), blank screen (200 ms), target stimulus (30 ms), mask (250 ms) and a final blank screen that remained until the subject gave a response. The next trial began after a variable interval between 500 and 1000 ms. For each trial, subjects responded by pressing one of two keys on a keyboard: "1" with the index finger for "bodies", or "2" with the middle finger for "chair" or *vice versa* (stimulus-response mapping was counterbalanced across participants). Every 32 trials subjects could take a break. Two blocks of familiarization preceded the experiment. In the first block, stimuli (four per condition) were shown for 250 ms, so that the subjects could easily see them. The second block (eight trials per condition) was identical to the actual experiment. Instructions for the familiarization blocks were identical to those of the actual experiment. The experiment lasted ~40 min. Stimulus presentation and response collection (accuracy and RTs) were controlled with Psychtoolbox (Brainard, 1997) through MATLAB. Each participant took part in two sessions of the same task. The second session took place with a delay ranging from two weeks to one month after the first one (mean interval 19.8 days ± 6.1 *SD*).

Autistic-Spectrum Quotient measure. Subjects completed the 50-items Autism-Spectrum Quotient scale (AQ; Baron-Cohen et al., 2011) after the second behavioral session, which measures the degree to which an adult with normal intelligence shows traits associated with the autism spectrum, in the domain of social skills, attention-switching, attention to detail, communication, and imagination. Subjects responded to each item using a 4-point rating scale, ranging from "definitely agree" to "definitely disagree".

fMRI data acquisition. Imaging was conducted on a MAGNETOM Prisma 3T scanner (Siemens Healthcare). T2*-weighted functional volumes were acquired using a gradient-echo echo-planar imaging sequence (GRE-EPI; TR/TE = 2000/30 ms, flip angle = 80°, acquisition matrix = 96 x 92, FOV = 210 x 201, 56 transverse slices, slice thickness = 2.2mm, no gap, multiband acceleration factor = 2 and phase encoding set to anterior/posterior direction). For the main experiment and the functional localizer task, we acquired eight runs for a total of 1546 frames per participant. Acquisition of high-resolution T1-weighted anatomical images was performed after the third functional run of the main experiment and lasted 8 min (MPRAGE; TR/TE/TI = 3000/3.7/1100 ms, flip angle = 8°, acquisition matrix = 320 x 280, FOV = 256 x 224 mm, slice thickness = 0.8mm, 224 sagittal slices, GRAPPA

accelerator factor = 2). Acquisition of two field maps was performed at the beginning of the fMRI session.

Main fMRI experiment. The fMRI data acquisition consisted of two parts: the main fMRI experiment, and a functional localizer task, which we describe below. In the main experiment, facing and non-facing dyads in upright or inverted orientation were presented over three runs, each lasting 6.83 min. Each run consisted of two sequences of 16 blocks (four blocks by condition), for a total of 32 blocks of eight second each. Blocks in the first sequence were presented in a random order, and blocks in the second sequence were presented in the counterbalanced (i.e., reversed) order relative to the first sequence. Each block featured eight images of the same condition, presented in a random order. A stimulus appeared once in a block, and twice in a run (once in each sequence). Thus, each of the three runs included eight blocks per condition (24 across the whole experiment). Three additional runs of 6.83 min each featured images of single bodies and single objects presented upright or inverted. Each run included 16 blocks (eight seconds each), eight with bodies and eight with objects (four with chairs and four with machines). Each block included eight different stimuli of the same condition. Each stimulus appeared twice in a run (once in each sequence). Runs with dyads and runs with single items (bodies or objects) were presented in pseudorandom order to avoid the occurrence of more than two consecutive runs of the same stimulus group. Each run began with a warm-up block (10 s) and ended with a cool-down block (16 s), during which a central fixation cross was presented. Within a run, the onset time of each block was jittered (range of inter-block interval duration: 2-6 s; total inter-block time by runs: 128 s), using the optseq tool of Freesurfer (Fischl, 2012) for optimal jittering. During each block, a black cross was always present in the center of the screen, while stimuli appeared for 550 ms, separated by an interval of 450 ms. In the 37% of all the stimulation and fixation periods, the cross changed color (from black to red). Participants were instructed to fixate the cross throughout the experiment, detect and report the color change by pressing a button with their right index finger. This task was used to minimize eye movements and maintain vigilance in the scanner. The main experiment lasted 41 min. During fMRI, stimuli were back-projected onto a screen by a liquid crystal projector (frame rate: 60 Hz; screen resolution: 1024x768 pixels, screen size: 40x30 cm). For all the stimuli, the center of the image overlapped with the center of the screen. Subjects, lying down inside the scanner, viewed the stimuli binocularly (~7° of visual angle) through a mirror above their head. Stimulus presentation, response collection and synchronization with the scanner were controlled with the Psychtoolbox (Brainard, 1997) through MATLAB (The MathWorks Inc, Natick, Massachusetts).

Functional localizer task. In addition to the six experimental runs, subjects completed a functional localizer task, with stimuli and task adapted from the fLoc package (Stigliani et al., 2015). During this task, subjects saw 180 grayscale photographs of the following five object classes: 1) body-stimuli (headless bodies in various views and poses, and body parts); 2) faces (adults and children); 3) places (houses and corridors); 4) inanimate objects (various exemplars of cars and musical instruments); 5) scrambled objects. Stimuli were presented over two runs (5.27 min each). Each run began with a warm-up block (12 s) and ended with a cool-down block (16 s), and included 72 blocks of four seconds each: 12 blocks for each object class with eight images per block (500 ms per image without interruption), randomly interleaved with 12 baseline blocks featuring an empty screen. To minimize low-level differences across categories, the view, size, and retinal position of the images varied across trials, and each item was overlaid on a 10.5° phase-scrambled background generated from another image of the set, randomly selected. During blocks containing images, some images were repeated twice, interleaved by a different image (2-back task). Participants had to press a button when they detected the repetition.

Preprocessing of fMRI data. Functional images were preprocessed and analyzed using MATLAB (MathWorks), in combination with SPM12 (Friston et al., 2007), the CoSMoMVPA toolbox (Oosterhof

et al., 2016), the MarsBaR toolbox (Brett et al., 2002) and the MatlabTFCE toolbox (Thornton, 2016). The first four volumes of each run were discarded, taking into account initial scanner gradient stabilization (Soares et al., 2016). Preprocessing of the remaining volumes involved despiking, slice time correction, geometric distortions correction using field maps, spatial realignment and motion correction using the first volume of each run as reference. The maximum displacement after despiking was 0.74 mm on the *x*-axis ($mean_{max} = 0.30$, $SD = 0.18$), 2.31 mm on the *y*-axis ($mean_{max} = 0.65$, $SD = 0.52$) and 3.29 mm on the *z*-axis ($mean_{max} = 1.06$, $SD = 0.64$), which did not exceed the Gaussian kernel of 5 mm FWHM used for the spatial smoothing before estimating the realignment parameters (Friston et al., 1996). Anatomical volumes were co-registered to the mean functional image, segmented into gray matter, white matter and cerebrospinal fluid in native space, and aligned to the probability maps in the Montreal Neurological Institute (MNI). The DARTEL method (Ashburner, 2007) was used to create a flow field for each subject and an inter-subject template, which was registered in the MNI space and used for normalization of functional images. Final steps included spatial smoothing with a Gaussian kernel of 6 mm FWHM and removing low-frequency drifts with a temporal high-pass filter (cutoff 128 s).

Analyses

Visual categorization task. Separately for the two sessions of the task, we computed the mean proportion of correct responses (accuracy) and the mean RTs for each condition, for each subject. Mean RTs were computed over trials with accurate response and RTs within 2 *SD* from the individual mean (discarded trials in session 1: 13.10% ±7.21 *SD*; in session 2: 10.20% ±5.87 *SD*). Accuracy and RT data were analyzed with 2 Category (body, chair) x 3 Stimulus (facing dyads, non-facing dyads, single bodies) x 2 Orientation (upright, inverted) repeated-measures ANOVAs. Critical comparisons were tested using pairwise *t* tests (two-tail). Analyses were repeated using average performance values from the two sessions, and for each session separately.

Test-retest reliability. For each subject, we computed the 2BIE on accuracy values [$(\text{Upright-inverted})_{facing}$-$(\text{Upright-inverted})_{non\text{-}facing}$] for each of the two sessions and compared them (*t* test, $\alpha = 0.05$, two-tail), to test for difference in the magnitude of the effect between sessions. Second, we computed the within-subjects Pearson correlation ($\alpha = 0.05$) between the 2BIE values in the first and second session, to assess the intra-individual consistency of the 2BIE. We ran a permutation test (10000 iterations) to assess whether the within-subjects correlation was higher than the correlation between any two randomly selected values (between-subjects correlations). To do so, in each iteration, a subject's 2BIE from the first session was randomly assigned to the 2BIE of another subject in the second session and a Pearson correlation coefficient was computed. We ranked the single within-subjects correlation value on the distribution of the 10000 between-subjects correlations values, and divided the rank by number of permutations to obtain a *p* value. We assessed whether a significant within-subject correlation was statistically higher than random between-subject correlations.

Correlation between 2BIE and AQ. We computed Spearman correlation ($\alpha = 0.05$) between the behavioral 2BIE and the AQ score of the 22 subjects who participated in both behavioral sessions. This analysis was repeated using data from session 1 of the visual categorization task (*Results*), from session 2 of the task and the average performance values from the two sessions (*Supplementary information*).

Definition of regions of interest (ROIs). Individual data recorded during the functional localizer task were entered into a General Linear Model with five regressors for the five object-class conditions (bodies, faces, places, objects and scrambled objects), one regressor for baseline blocks, and six regressors for movement correction parameters as nuisance covariates. Three bilateral masks of the middle temporo-occipital gyrus (MTOG), the temporo-occipital fusiform cortex (TOFC) and the inferior

parahippocampal cortex (PHC) were created using FSLeyes (McCarthy, 2019) and the Harvard-Oxford Atlas (Desikan et al., 2006) through FSL (Jenkinson et al., 2012). In addition, a bilateral mask of the EVC encompassing visual areas V1/V2 and V3v/V3d, was created using the SPM Anatomy toolbox (Eickhoff et al., 2005). Within each mask of each subject, we selected the voxels with significant activity (threshold: $p = 0.05$) for the contrasts of interest, to create the ROIs: EBA in the MTOG with the contrast bodies > [objects+faces+places], FFA in the TOFC with the contrast faces > [objects+bodies+places], FBA in the TOFC with the contrast body > [objects+face+places] and PPA in the PHC with the contrast places > [objects+faces+bodies]. The EVC-ROI included all the voxels responsive to visual stimuli (contrast: [bodies+faces+places+objects+scrambled objects] > baseline). For each ROI, all of the voxels within the bilateral mask that passed the threshold were ranked by activation level based on *t* values. The final ROIs included up to 200 best voxels across the right and left ROI (number of voxels in the EBA, FBA, FFA and EVC: 200; in the PPA: 199.86 ± 0.58 *SD*). Neural activity values used for all ROIs analyses were extracted as the percent signal change (PSC) for each condition, using the MarsBaR toolbox (Brett et. al., 2002). When a functional ROI could not be identified (the FBA for one subject and all five ROIs for another, due to a technical failure during the functional localizer task), PSC were extracted from a mask corresponding to the group-level activation for the contrast of interest.

ROIs analyses. Using data recorded during dyad runs of the main fMRI experiment (N = 29), we performed a repeated-measures ANOVA with factors: 5 ROI (EBA, FBA, FFA, PPA and EVC) x 2 Stimulus (facing and non-facing dyads) x 2 Orientation (upright and inverted). Effects in each ROI were further inspected using pairwise *t* tests to address the difference between upright facing *vs.* non-facing dyads ($\alpha = 0.05$ one-tail) and difference in the inversion effect (upright > inverted) between facing and non-facing dyads ($\alpha = 0.05$, one-tail). For each ROI, data recorded during runs with single bodies and objects were first analyzed separately with a 2 Stimulus x 2 Orientations repeated-measures ANOVA, followed up with pairwise *t* tests, to assess the existence of an inversion effect separately for bodies and objects ($\alpha = 0.05$, one-tail) and the difference in the inversion effect (upright > inverted) between bodies and objects ($\alpha = 0.05$, two-tail). Second, from each ROI, we extracted multivariate activity patterns for the four conditions, and performed MVPA to compute the accuracy of discrimination between upright and inverted stimuli, for bodies and objects. To do so, a support vector machine classifier was trained on 5 out of 6 runs to classify upright *vs.* inverted stimuli, and then tested on data from the remaining run. This scheme was iterated until all runs were used as testing and training sets. For each subject, for each ROI, accuracy values were averaged across all iterations and tested against chance (50% accuracy, *t* tests, $\alpha = 0.05$, one-tail) and, finally, compared between bodies and objects (*t* tests, $\alpha = 0.05$, two-tail).

Brain-behavior correlation (in ROIs). For the 22 subjects who were included in all behavioral and fMRI sessions, we computed the (normalized) neural 2BIE, using the PSC extracted from each ROI for each condition (upright and inverted facing and non-facing dyads). For each ROI, the values representing the individual neural 2BIEs were correlated with the values representing the individual behavioral 2BIE (Pearson correlation, $\alpha = 0.05$). This analysis was repeated using data from session 1 of the behavioral task (*Results*), from session 2 and the average performance across the two sessions (*Supplementary Information*).

Acknowledgements

This work was supported by a European Research Council Starting Grant to L.P. (Grant number: THEMPO-758473).

We thank Nicolas Goupil for drafting an earlier version of Figure 1.

References

Abassi E, Papeo L (2020) The representation of two-body shapes in the human visual cortex. J Neurosci 40:852–863. https://doi.org/10.1523/JNEUROSCI.1378-19.2019

Adibpour P, Hochmann J-R, Papeo L (2021) Spatial Relations Trigger Visual Binding of People. J Cogn Neurosci 33:1343–1353. https://doi.org/10.1162/jocn_a_01724

Ashburner J (2007) A fast diffeomorphic image registration algorithm. Neuroimage 38:95–113. https://doi.org/10.1016/j.neuroimage.2007.07.007

Atkinson, A. P., Heberlein, A. S., Adolphs, R., Adams, R. B., Ambady, N., Nakayama, K., & Shimojo, S. (2011). Are people special? A brain's eye view. The science of social vision, 363-392. https://doi.org/10.1093/acprof:oso/9780195333176.003.0022

Baron-Cohen, S. (1997). How to build a baby that can read minds: Cognitive mechanisms in mindreading. The maladapted mind: Classic readings in evolutionary psychopathology, 207-239.

Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E (2001) The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. J Autism Dev Disord 31:5–17. https://doi.org/10.1023/A:1005653411471

Bate S, Frowd C, Bennetts R, Hasshim N, Portch E, Murray E, Dudfield G (2019) The consistency of superior face recognition skills in police officers. Appl Cogn Psychol 33:828–842. https://doi.org/10.1002/acp.3525

Bellot E, Abassi E, Papeo L (2021) Moving Toward versus Away from Another: How Body Motion Direction Changes the Representation of Bodies and Actions in the Visual Cortex. Cereb Cortex 31:2670–2685. https://doi.org/10.1093/cercor/bhaa382

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436. https://doi.org/10.1163/156856897X00357

Brambilla P, Hardan A, Ucelli Di Nemi S, Perez J, Soares JC, Barale F (2003) Brain anatomy and development in autism: Review of structural MRI studies. Brain Res Bull 61:557–569. https://doi.org/10.1016/j.brainresbull.2003.06.001

Brandman T, Yovel G (2010) The Body Inversion Effect Is Mediated by Face-Selective, Not Body-Selective, Mechanisms. 30:10534–10540. https://doi.org/10.1523/JNEUROSCI.0911-10.2010

Brandman T, Yovel G (2012) A face inversion effect without a face. Cognition 125:365–372. https://doi.org/10.1016/j.cognition.2012.08.001

Brandman T, Yovel G (2016) Bodies are Represented as Wholes Rather Than Their Sum of Parts in the Occipital-Temporal Cortex. 26(2):530–543. https://doi.org/10.1093/cercor/bhu205

Brett M, Anton JL, Valabregue R, Poline JB (2002) Region of interest analysis using a SPM toolbox. 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan.

Cook J, Barbalat G, Blakemore SJ (2012) Top-down modulation of the perception of other people in schizophrenia and autism. Front Hum Neurosci 6:1–10. https://doi.org/10.3389/fnhum.2012.00175

Dawson G, Webb SJ, McPartland J (2005) Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies. Dev Neuropsychol 27:403–424. https://doi.org/10.1207/s15326942dn2703_6

Dennett HW, McKone E, Edwards M, Susilo T (2012) Face Aftereffects Predict Individual Differences in Face

Recognition Ability. Psychol Sci 23:1279–1287. https://doi.org/10.1177/0956797612446350

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31:968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Di Giorgio E, Frasnelli E, Rosa Salva O, Maria Luisa S, Puopolo M, Tosoni D, Simion F, Vallortigara G, Apicella F, Gagliano A, Guzzetta A, Molteni M, Persico A, Pioggia G, Valeri G, Vicari S (2016) Difference in Visual Social Predispositions between Newborns at Low-and High-risk for Autism. Sci Rep 6:1–9. https://doi.org/10.1038/srep26395

Duchaine B, Nakayama K (2006) The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. Neuropsychologia 44:576–585. https://doi.org/10.1016/j.neuropsychologia.2005.07.001

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25:1325–1335. https://doi.org/10.1016/j.neuroimage.2004.12.034

Elliott ML, Knodt AR, Ireland D, Morris ML, Poulton R, Ramrakha S, Sison ML, Moffitt TE, Caspi A, Hariri AR (2020) What Is the Test-Retest Reliability of Common Task-Functional MRI Measures? New Empirical Evidence and a Meta-Analysis. Psychol Sci 31:792–806. https://doi.org/10.1177/0956797620916786

Farroni, T., & Senju, A. (2011). Specialized brain for the social vision: Perspectives from typical and atypical development. Social Vision, 421-443. https://doi.org/10.1093/acprof:oso/9780195333176.003.0025

Fischl B (2012) FreeSurfer. Neuroimage 62:774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Friston KJ, Williams S, Howard R, Frackowiak RSJ, Turner R (1996) Movement-Related effects in fMRI time-series. Magn Reson Med 35:346–355. https://doi.org/10.1002/mrm.1910350312

Friston KJ, Ashburner J, Kiebel S, Nichols T, Penny WD (2007) Statistical parametric mapping : the analysis of funtional brain images. Elsevier/Academic Press.

Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC (1999) Activation of the middle fusiform ‘ face area ’ increases with expertise in recognizing novel objects. Nat Neuroscience 2(6), 568-573. https://doi.org/10.1038/9224

Gauthier I, Tarr MJ, Moylan J, Skudlarski P, Gore JC, Anderson AW (2000) The fusiform “face area” is part of a network that processes faces at the individual level. J Cogn Neurosci 12:495–504. https://doi.org/10.1162/089892900562165

Hadad BS, Schwartz S, Binur N (2019) Reduced perceptual specialization in autism: Evidence from the other-race face effect. J Exp Psychol Gen 148:588–594. https://doi.org/10.1037/xge0000550

Hafri A, Firestone C (2021) The Perception of Relations. Trends Cogn Sci:1–18. https://doi.org/10.1016/j.tics.2021.01.006

Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012) FSL. Neuroimage 62:782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Johnson CP (2004) New tool helps primary care physicians diagnose autism early. AAP News, 24(2), 74. https://www.aappublications.org/content/24/2/74

Kaiser D, Quek GL, Cichy RM, Peelen M V. (2019) Object Vision in a Structured World. Trends Cogn Sci 23:672–685. https://doi.org/10.1016/j.tics.2019.04.013

Lander K, Bruce V, Bindemann M (2018) Use-inspired basic research on individual differences in face identification: implications for criminal investigation and security. Cogn Res Princ Implic 3:1–13. https://doi.org/10.1186/s41235-018-0115-6

Laycock R, Wood K, Wright A, Crewther SG, Goodale MA (2020) Saccade Latency Provides Evidence for Reduced Face Inversion Effects With Higher Autism Traits. Front Hum Neurosci 13. https://doi.org/10.3389/fnhum.2019.00470

Maurer D, Le Grand R, Mondloch CJ (2002) The many faces of configural processing. Trends Cogn Sci 6:255–260. https://doi.org/10.1016/s1364-6613(02)01903-4

McCarthy P (2018) FSLeyes. https://doi.org/10.5281/zenodo.1470761

Neuhaus E, Jones EJH, Barnes K, Sterling L, Estes A, Munson J, Dawson G, Webb SJ (2016) The Relationship Between Early Neural Responses to Emotional Faces at Age 3 and Later Autism and Anxiety Symptoms in Adolescents with Autism. J Autism Dev Disord 46:2450–2463. https://doi.org/10.1007/s10803-016-2780-y

Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: A primer with examples. Hum Brain Mapp 15(1):1–25. https://doi.org/10.1002/hbm.1058

Oosterhof NN, Connolly AC, Haxby J V. (2016) CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. Front Neuroinform 10:1–27. https://doi.org/10.3389/fninf.2016.00027

Orlandi A, Proverbio AM (2020) ERP indices of an orientation-dependent recognition of the human body schema. Neuropsychologia 146:107535. https://doi.org/10.1016/j.neuropsychologia.2020.107535

Papeo L (2020) Twos in human visual perception. Cortex 132:473-478. https://doi.org/10.1016/j.cortex.2020.06.005

Papeo L, Abassi E (2019) Seeing social events: The visual specialization for dyadic human–human interactions. J Exp Psychol Hum Percept Perform 45:877–888. https://doi.org/10.1037/xhp0000646

Papeo L, Stein T, Soto-Faraco S (2017) The Two-Body Inversion Effect. Psychol Sci 28:369–379. https://doi.org/10.1177/0956797616685769

Peelen M V., Wiggett AJ, Downing PE (2006) Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. Neuron 49:815–822. https://doi.org/10.1016/j.neuron.2006.02.004

Peelen M V, Downing PE (2007) The neural basis of visual body perception. 8:636–648. https://doi.org/10.1038/nrn2195

Powell LJ, Spelke ES (2018) Human infants' understanding of social imitation: Inferences of affiliation from third party observations. Cognition 170:31–48. https://doi.org/10.1016/j.cognition.2017.09.007

Quadflieg S, Koldewyn K (2017) The neuroscience of people watching: How the human brain makes sense of other people's encounters. Ann N Y Acad Sci 1396:166–182. https://doi.org/10.1111/nyas.13331

Ramon M, Bobak AK, White D (2019) Super-recognizers: From the lab to the world and back again. Br J Psychol 110:461–479. https://doi.org/10.1111/bjop.12368

Reed CL, Stone VE, Bozova S, Tanaka J (2003) The body-inversion effect. Psychol Sci 14:302–308. https://doi.org/10.1111/1467-9280.14431

Reed CL, Stone VE, Grubb JD, McGoldrick JE (2006) Turning configural processing upside down: Part and whole body postures. J Exp Psychol Hum Percept Perform 32:73–87. https://doi.org/10.1037/0096-1523.32.1.73

Rezlescu C, Chapman A, Susilo T, Caramazza A (2016). Large inversion effects are not specific to faces and do not vary with object expertise. PsyArXiv. https://doi.org/10.31234/osf.io/xzbe5

Rezlescu C, Wilmer JB, Caramazza A (2017) The Inversion , Part-Whole , and Composite Effects Reflect Distinct Perceptual Mechanisms With Varied Relationships to Face Recognition. J Exp Psychol Hum Percept Perform 43(12):1961. https://doi.org/10.1037/xhp0000400

Rhodes G, Byatt G, Michie PT, Puce A (2004) Is the Fusiform Face Area Specialized for Faces, Individuation, or Expert Individuation? J Cogn Neurosci 16:189–203. https://doi.org/10.1162/089892904322984508

Rule NO, Ambady N, Adams RB (2009) Personality in perspective: Judgmental consistency across orientations of the face. Perception 38:1688–1699. https://doi.org/10.1068/p6384

Ruzich, Emily Allison, Carrie Smith P, Watson P et al. (2015) Measuring autistic traits in the general population: a systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. Mol Autism 6:1–12. https://doi.org/10.1186/2040-2392-6-2

Simmons DR, Robertson AE, McKay LS, Toal E, McAleer P, Pollick FE (2009) Vision in autism spectrum disorders. Vision Res 49:2705–2739. https://doi.org/10.1016/j.visres.2009.08.005

Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44:83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Soares JM, Magalhães R, Moreira PS, Sousa A, Ganz E, Sampaio A, Alves V, Marques P, Sousa N (2016) A Hitchhiker's Guide to Functional Magnetic Resonance Imaging. Front Neurosci 10:515. https://doi.org/10.3389/fnins.2016.00515

Stekelenburg JJ, De Gelder B (2004) The neural correlates of perceiving human bodies: An ERP study on the body-inversion effect. Neuroreport 15:777–780. https://doi.org/10.1097/00001756-200404090-00007

Stigliani A, Weiner KS, Grill-Spector K (2015) Temporal Processing Capacity in High-Level Visual Cortex Is Domain Specific. J Neurosci 35:12412–12424. https://doi.org/10.1523/jneurosci.4822-14.2015

Susilo T, Germine L, Duchaine B (2013) Face recognition ability matures late: Evidence from individual differences in young adults. J Exp Psychol Hum Percept Perform 39:1212–1217. https://doi.org/10.1037/a0033469

Tanaka JW, Farah MJ (1993) Parts and Wholes in Face Recognition. Q J Exp Psychol Sect A 46:225–245. https://doi.org/10.1080/14640749308401045

Taylor JC, Wiggett AJ, Downing PE (2007) Functional MRI Analysis of Body and Body Part Representations in the Extrastriate and Fusiform Body Areas. J Neurophysiol 98:1626–1633. https://doi.org/10.1152/jn.00012.2007

Thornton (2016). Introducing MatlabTFCE. http://markallenthornton.com/blog/matlab-tfce/ https://github.com/markallenthornton/MatlabTFCE

Troje NF, Westhoff C (2006) The Inversion Effect in Biological Motion Perception: Evidence for a "Life Detector"? Curr Biol 16:821–824. https://doi.org/10.1016/j.cub.2006.03.022

van Boxtel JJA, Peng Y, Su J, Lu H (2017) Individual differences in high-level biological motion tasks correlate with autistic traits. Vision Res 141:136–144. https://doi.org/10.1016/j.visres.2016.11.005

Van der Donck S, Dzhelyova M, Vettori S, Thielen H, Steyaert J, Rossion B, Boets B (2019) Fast Periodic Visual Stimulation EEG Reveals Reduced Neural Sensitivity to Fearful Faces in Children with Autism. J Autism Dev Disord 49:4658–4673. https://doi.org/10.1007/s10803-019-04172-0

Walbrin J, Koldewyn K (2019) Dyadic interaction processing in the posterior temporal cortex. Neuroimage 198:296–302. https://doi.org/10.1016/j.neuroimage.2019.05.027

Wang R, Li J, Fang H, Tian M, Liu J (2012) Individual differences in holistic processing predict face recognition ability. Psychol Sci 23:169–177. https://doi.org/10.1177/0956797611420575

Webb SJ, Neuhaus E, Faja S (2017) Face perception and learning in autism spectrum disorders. Q J Exp Psychol 70:970–986. https://doi.org/10.1080/17470218.2016.1151059

Wyer NA, Martin D, Pickup T, Neil Macrae C (2012) Individual Differences in (Non-Visual) Processing Style Predict the Face Inversion Effect. Cogn Sci 36:373–384. https://doi.org/10.1111/j.1551-6709.2011.01224.x

Yin RK (1969) Looking at upside-down faces. J Exp Psychol 81:141–145. https://psycnet.apa.org/doi/10.1037/h0027474

Yovel G, Kanwisher N (2004) Face Perception : Domain Specific , Not Process Specific. Neuron 44(5):889–898. https://doi.org/10.1016/j.neuron.2004.11.018

Yovel G, Kanwisher N (2005) The neural basis of the behavioral face-inversion effect. Curr Biol 15:2256–2262. https://doi.org/10.1016/j.cub.2005.10.072

Yovel G, Wilmer JB, Duchaine B (2014) What can individual differences reveal about face processing? Front Hum Neurosci 8:1–9. https://doi.org/10.3389/fnhum.2014.00562